

# INSTITUTE FOR PHYSICAL SCIENCE AND TECHNOLOGY

AD A053026

University of  
Maryland  
College Park



12

Technical Note BN-869

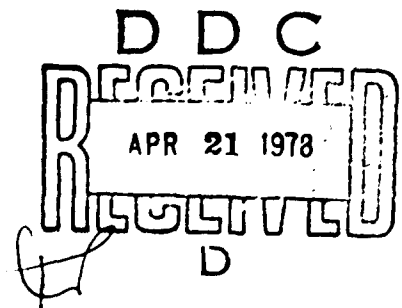
## ANALYSIS OF OPTIMAL FINITE ELEMENT MESHES IN $R^1$

by

I. Babuska

and

W. C. Rheinboldt



March 1978

**DISTRIBUTION STATEMENT A**

Approved for public release;  
Distribution Unlimited



### Abstract

A theory of a-posteriori estimates for the finite element method was developed earlier by the authors. Based on this theory, for a two-point boundary value problem the existence of a unique optimal mesh distribution is proved and its properties analyzed. This mesh is characterized in terms of certain, easily computable local error indicators which in turn allow for a simple adaptive construction of the mesh and also permit the computation of a very effective a-posteriori error bound. While the error estimates are asymptotic in nature, numerical experiments show the results to be excellent already for 10% accuracy. The approaches are not restricted to the model problem considered here only for clarity; in fact, they allow for rather straightforward extensions to more general problems in one dimension as well as to higher order elements.

## 1. Introduction

For the numerical solution of boundary value problems by finite-element techniques, the construction of optimal, or near-optimal meshes is of considerable practical importance. The same can be said when finite-difference or collocation methods are used. Many articles in the literature deal with questions that bear a relation to this problem, yet, as observed in [15], even for two-point boundary value problems relatively few address it directly. We shall not attempt to survey this literature.

There are various analyses of the approximation error of a given function by piecewise polynomials with a fixed number of pieces of fixed order (see, e.g., [7], [8], [10], [12], [21] and the references cited there). In principle, such studies may relate to the finite element method since that method leads to optimal approximations under the energy norm. The mentioned error estimates involve higher derivatives of the given function. With these results as a basis, a number of authors developed methods for the construction of optimal meshes for collocation and finite-difference methods ([11], [13], [14], [17], [22], [23]). For this the needed information about the derivatives of the solution is obtained from the approximate solution, for instance, by means of difference formulas. This procedure can be theoretically justified in the case of regular meshes (see, e.g., [20]). However, when there are abrupt changes in the mesh--as they arise with refinement techniques--then "internal boundary layers" appear in the error function (see, e.g., [2], [6]), and hence the difference formulas cease to approximate well the desired derivatives.

Various results on optimizing finite element meshes have appeared in the engineering literature. Without entering into any details, we mention, for instance, the articles [9], [18], [19], [25], [26], and [27].

In recent year, for initial value problems for ordinary differential equations very effective procedures have been designed and analyzed for adapting the stepsizes and the order of the numerical methods (see, e.g., the survey [15]). The principal tool is the availability of an error analysis with a local, a-posteriori character. These estimates are asymptotic in nature; yet practical experience has proved their reliability for reasonable tolerances.

In this paper we use a new approach to the construction of optimal finite element meshes. It is based on a theory of a-posteriori estimates for the finite element method developed in [3], [4] (see also [5]). As in the case of the initial value problems, the estimates are asymptotic in character. More specifically, higher-order terms in the maximal mesh-size  $\bar{h}$  are neglected; that is, asymptotic expressions of the form  $1 + o(1)$  as  $\bar{h} \rightarrow 0$  are considered to be approximately equal to one. At the same time, all constant factors of these  $(1+o(1))$ -terms can be evaluated computationally.

For clarity of presentation, we restrict the discussion to a simple two-point boundary value problem involving a linear, self-adjoint operator of second order. Moreover, for simplicity, we employ only piecewise linear elements. The approaches allow for rather straightforward extensions to a variety of more general problems in one dimension, and there are no essential limitations to the use of higher-order elements. In fact, analogous

techniques even permit consideration of elements of different order in different parts of the mesh.

Continuous mesh distribution functions are used to prove the existence of a unique, optimal mesh distribution and to analyze its properties. In particular, it is shown that the value of the optimal error is rather stable under perturbations of the optimal mesh. Hence it is indeed unnecessary to compute this mesh with excessive accuracy. The optimal mesh is characterized in terms of certain easily computable local error indicators. This allows for a simple adaptive method to construct that mesh (see [4]) and, at the same time, to compute very effective a-posteriori error bounds. Although, as mentioned, the error estimates have only asymptotic character, numerical experience shows that, as in the case of initial value problems, the results are excellent already for accuracies of the order of ten percent.

## 2. Notation

Let  $I = I(\alpha, \beta)$ ,  $\alpha < \beta$ , be the open interval  $\{x \in \mathbb{R}^1; \alpha < x < \beta\}$  and  $\bar{I} = \bar{I}(\alpha, \beta)$  its closure. As usual,  $H^0(I)$  denotes the space of square-integrable functions on  $\bar{I}$  and  $C^0(I) \subset H^0(I)$  the subspace of continuous functions on  $\bar{I}$ . The norms on  $H^0(I)$  and  $C^0(I)$  will be written as  ${}_I\|\cdot\|_0$  and  ${}_I\|\cdot\|_c$ , respectively.

Define  $E(I)$  as the space of real, infinitely differentiable functions on  $I$  for which all derivatives have continuous extensions on  $\bar{I}$ . Moreover, let  $\mathcal{D}(I) \subset E(I)$  be the subspace of all functions with compact support in  $I$ . For any integer  $k \geq 1$ , the spaces  $H^k(I)$  and  $C^k(I)$  are the completions of  $E(I)$  under the norms

$$(2.1) \quad {}_I\|u\|_k^2 = \sum_{i=0}^k {}_I\left\|\frac{d^i u}{dx^i}\right\|_0^2$$

and

$$(2.2) \quad {}_I\|u\|_{c,k} = \sum_{i=0}^k {}_I\left\|\frac{d^i u}{dx^i}\right\|_c,$$

respectively. Analogously, the completions of  $\mathcal{D}(I)$  under these two norms are the spaces  $H_0^k(I)$  and  $C_0^k(I)$ .

Let  $a, b \in C^0(I)$  be given such that  $a(x) \geq \underline{a} > 0$ ,  $b(x) \geq 0$ ,  $\forall x \in I$ . Then  $E(I)$  and  $E_0(I)$  shall be the spaces  $H^1(I)$  and  $H_0^1(I)$ , respectively, with their norm replaced by

$$(2.3) \quad \|u\|_E^2 = \left[ \int_I (au'^2 + bu^2) dx \right], \quad (u' = \frac{du}{dx}).$$

If  $b \equiv 0$  on  $I$ , then (2.3) is only a seminorm on  $E(I)$ . On the other hand, on  $E_0(I)$ , (2.3) is always a norm which, moreover, is equivalent to  $\| \cdot \|_1$ . Obviously,  $E_0(I)$  is a Hilbert space and for  $b \neq 0$  the same is true for  $E(I)$ . We denote the inner product in either space by  $(\cdot, \cdot)_E$ . For  $b \equiv 0$  on  $I$ ,  $E(I)$  is a Hilbert space modulo the constant functions.

On  $\bar{I} = \bar{I}(\alpha, \beta)$  we consider partitions

$$(2.4) \quad \Delta(\cdot) : \alpha = x_0^\Delta < x_1^\Delta < \dots < x_{m-1}^\Delta < x_m^\Delta = \beta, \quad m = m(\Delta) \geq 1,$$

and introduce the notations

$$(2.5) \quad \left. \begin{aligned} I_j(\Delta) &= I(x_{j-1}^\Delta, x_j^\Delta) \\ h_j(\Delta) &= x_j^\Delta - x_{j-1}^\Delta \end{aligned} \right\} j = 1, 2, \dots, m$$

$$\bar{h}(\Delta) = \max_{j=1, \dots, m} h_j(\Delta), \quad \underline{h}(\Delta) = \min_{j=1, \dots, m} h_j(\Delta)$$

All partitions  $\Delta$  which for fixed  $\lambda > 0$ ,  $\kappa \geq 1$  satisfy

$$(2.6) \quad \underline{h}(\Delta) \geq \lambda \bar{h}(\Delta)^\kappa$$

are said to be  $(\lambda, \kappa)$ -regular.

For given  $\Delta = \Delta(I)$ , we denote by  $S(I, \Delta) \subset H^1(I)$  and  $S_0(I, \Delta) \subset H_0^1(I)$  the subspaces of all functions for which the restriction to any  $I_j(\Delta)$ ,  $j = 1, \dots, m$ , is linear. Analogously  $P^k(I, \Delta) \subset H^1(I)$  and  $P_0^k(I, \Delta) \subset H_0^1(I)$ ,  $k \geq 0$ , shall consist of the functions for which the restrictions to  $I_j(\Delta)$ ,  $j = 1, \dots, m$ , belong to  $C^k(I_j)$ .



For later use we note the following well-known lemma (see, e.g., [24]):

Lemma 2.1: For given  $I = I(\alpha, \beta)$ ,  $\alpha < \beta$ , and  $\Delta = \Delta(I)$  there exists a positive constant  $K$  such that

$$(2.7) \quad \inf_{w \in S_0(I, \Delta)} \|u - w\|_E \leq K \tilde{h}(\Delta)_I \|u\|_2, \quad \forall u \in H^2(I) \cap H_0^1(I).$$

### 3. A Boundary Value Problem

#### 3.1 Basic Formulation

As mentioned in the introduction, we restrict the discussion to a simple model problem. For ease of notation, the unit interval  $I = I(0,1)$  is used from now on throughout the remainder of the paper. On  $I$  we consider the equation

$$(3.1) \quad L[u] \equiv - \frac{d}{dx} a(x) \frac{du}{dx} + b(x)u = f(x), \quad x \in I(0,1) ,$$

together with the boundary conditions

$$(3.2) \quad u(0) = u(1) = 0 .$$

We assume that  $a \in C^2(I)$ ,  $b, f \in C^1(I)$ , and, as before, that  $a(x) \geq \underline{a} > 0$ ,  $b(x) \geq 0$ ,  $\forall x \in I$ .

The weak solution of the problem is the unique  $u_0 \in E_0(I)$  with

$$(3.3) \quad {}_I(u_0, v)_E = F_f(v), \quad \forall v \in E_0(I)$$

where

$$(3.4) \quad F_f(v) = \int_I f v dx .$$

Note that under our differentiability assumptions about  $a, b, f$  the solution  $u_0$  of (3.3) belongs to  $C^3(I)$  and also satisfies (3.1/2).

With the partition  $\Delta$  and the space  $S_0(I, \Delta) \subset H_0^1(I)$  specified as in Section 2, we consider the finite element solution  $u_\Delta \in S_0(I, \Delta)$  defined by

$$(3.5) \quad {}_I(u_\Delta, v)_E = F_f(v), \quad \forall v \in S_0(I, \Delta).$$

Since  $u_0 \in H^2(I) \cap H_0^1(I)$ , it then follows from Lemma 2.1 that

$$(3.6) \quad {}_I\|u_\Delta - u_0\|_E \leq K \bar{h}(\Delta) {}_I\|u_0\|_2.$$

### 3.2 A-Posteriori Error Analysis

We consider the residual  $r = L(u_\Delta) - f$  on the intervals  $I_j$ , that is,

$$(3.7) \quad r_j(x) = (L(u_\Delta) - f)(x), \quad \forall x \in I_j, \quad j = 1, \dots, m.$$

Let  $z_j \in E_0(I_j)$  be the solutions of

$$(3.8) \quad {}_{I_j}(z_j, v)_E = F_{r_j}(v), \quad \forall v \in E_0(I_j), \quad j = 1, \dots, m$$

and set

$$(3.9) \quad Z(\Delta)^2 = \sum_{j=1}^m {}_{I_j}\|z_j\|_E^2.$$

The following result was proved in [3]:

Theorem 3.1: The error  $e = u_\Delta - u_0$  satisfies

$$(3.10) \quad {}_I\|e\|_E^2 = Z(\Delta)^2(1 + O(\bar{h}(\Delta))) \text{ as } \bar{h}(\Delta) \rightarrow 0,$$

where the constant in the bound of the ~~0-term~~ depends on  $a$  and  $b$  but not on  $f$  and  $\Delta$ .

We analyze the quantity  $Z(\Delta)$  of (3.9) in some more detail. Because  $u_\Delta$  is linear on any  $I_j$ , we may write

$$(3.11) \quad r_j = L(u_\Delta) - f = -a'u_\Delta' + bu_\Delta - f = \rho_j + \tau_j,$$

where

$$(3.12) \quad \rho_j(x) = a(x)u_0''(x), \tau_j(x) = -a'(x)e'(x) + b(x)e(x), \forall x \in I_j.$$

Let  $\varphi_j, \psi_j \in E_0(I_j)$  be such that

$$(3.13) \quad \begin{aligned} (a) \quad & I_j(\varphi_j, v)_E = F_{\rho_j}(v), \forall v \in E_0(I_j) \\ (b) \quad & I_j(\psi_j, v)_E = F_{\tau_j}(v), \forall v \in E_0(I_j), \end{aligned}$$

and therefore  $z_j = \varphi_j + \psi_j$ .

The smallest eigenvalue of the differential operator  $L$  on  $I_j$  with zero boundary conditions is bounded below by the smallest eigenvalue  $\pi^2/h_j^2$  of the operator  $-ad^2/dx^2$  on  $I_j$ . Hence it follows from (3.13b) that

$$I_j \|\psi_j\|_0 \leq Ch_j^2 I_j \|\tau_j\|_0$$

and therefore

$$(3.14) \quad I_j \|\psi_j\|_E^2 = F_{\tau_j}(\psi_j) \leq I_j \|\tau_j\|_0 I_j \|\psi_j\|_0 \leq Ch_j^2 I_j \|\tau_j\|_0^2.$$

Here as in subsequent estimates  $C$  denotes a generic constant which has different values in each instance but is independent of the other essential variables in the same expression. Now note that

$$(3.15) \quad I_j \|\tau_j\|_0^2 = \int_{I_j} (-a'e' + be)^2 dx \leq C \int_{I_j} [(a'e')^2 + (be)^2] dx \leq C I_j \|e\|_E^2,$$

which together with (3.14) gives

$$(3.16) \quad I_j \| \psi_j \|_E \leq C_{\alpha_j} I_j \| e \|_E .$$

We introduce the quantities

$$R(\Delta) = \frac{1}{I \| e \|_E} \left[ \sum_{j=1}^m I_j \| \psi_j \|_E^2 \right]^{1/2}$$

and

$$(3.17) \quad Q(\Delta)^2 = \sum_{j=1}^m I_j \| \psi_j \|_E^2 .$$

From Theorem 3.1 we obtain — with some  $|\alpha| \leq 1$  —

$$\begin{aligned} I \| e \|_E^2 &= Z^2(\Delta)(1+O(\bar{h})) = \sum_{j=1}^m I_j (\phi_j + \psi_j, \phi_j + \psi_j)_E (1+O(\bar{h})) \\ &= [Q(\Delta)^2 + 2\alpha Q(\Delta) I \| e \|_E R(\Delta) + I \| e \|_E^2 R(\Delta)^2] (1+O(\bar{h})) \\ &= (Q(\Delta) + \alpha I \| e \|_E R(\Delta))^2 (1+O(\bar{h})) + (1-\alpha^2) I \| e \|_E^2 R(\Delta)^2 (1+O(\bar{h})) . \end{aligned}$$

But by (3.16) we have  $R(\Delta) = O(\bar{h})$  and thus

$$I \| e \|_E^2 = (Q(\Delta) + \alpha I \| e \|_E R(\Delta))^2 (1+O(\bar{h})) ,$$

which in turn implies that

$$(3.20) \quad I \| e \|_E = Q(\Delta)(1+O(\bar{h})) .$$

Therefore, in view of (3.10), we have proved the following result:

Theorem 3.2: Let  $Z(\Delta)$  and  $Q(\Delta)$  be defined by (3.3) and (3.15), respectively. Then

$$(3.21) \quad Q(\Delta)^2 = Z(\Delta)^2(1+O(\bar{h})), \text{ as } \bar{h}(\Delta) \rightarrow 0,$$

where the constant in the bound of the  $O$ -term depends on  $a$  and  $b$  but not on  $f$  and  $\Delta$ .

While approximations of  $Z(\Delta)$  can be computed and  $Q(\Delta)$  is not readily accessible, the quantity  $Q$  is better suited than  $Z$  for our theoretical studies of optimal partitions  $\Delta$ .

#### 4. Optimal Partitions - Case I

In this section we restrict ourselves to the case when  $u_0'' \neq 0$  on  $I$ . This condition will be removed in Section 5.

##### 4.1 Representations of $Q(\Delta)$

Recall that under our assumptions about  $a, b, f$  we have  $u_0 \in C^3(I)$  and hence  $\rho = au_0'' \in C^1(I)$ .

Lemma 4.1: Suppose that  $u_0''(x) \neq 0$  for all  $x \in I$  and set  $x_{j-1/2} = (x_j + x_{j-1})/2$ ,  $a_{j-1/2} = a(x_{j-1/2})$ ,  $\bar{\rho}_j = \rho(\xi_j)$  where  $|\rho(\xi_j)| = \max\{|\rho(x)|, x \in I_j\}$ . Then

$$(4.1) \quad Q(\Delta)^2 = \frac{1}{12} \left[ \sum_{j=1}^m \frac{\bar{\rho}_j^2}{a_{j-1/2}} h_j^3 \right] (1 + O(\bar{h}(\Delta))), \text{ as } \bar{h}(\Delta) \rightarrow 0,$$

where the constant in the bound of the  $O$ -term depends on  $a, b$  and  $I \|f\|_{C,1}$ .

Proof: Set

$$(4.2) \quad \sigma_j(x) = \rho(x) - \bar{\rho}_j, \quad \forall x \in I_j, \quad j = 1, \dots, m$$

and define  $\phi_{1,j}, \phi_{2,j} \in E_0(I_j)$  as the solutions of

$$(4.3) \quad \left. \begin{aligned} I_j^{(\phi_{1,j}, v)_E} &= F_{\bar{\rho}_j}(v), \quad \forall v \in E_0(I_j) \\ I_j^{(\phi_{2,j}, v)_E} &= F_{\sigma_j}(v), \quad \forall v \in E_0(I_j) \end{aligned} \right\} \quad j = 1, \dots, m$$

respectively. Then we have  $\phi_j = \phi_{1j} + \phi_{2j}$ .

By assumption there exists a constant  $\rho_0 > 0$  such that  $|\rho(x)| \geq \rho_0$  for all  $x \in I$ , and hence that for all sufficiently small  $\bar{h}$

$$(4.4) \quad |\sigma_j(x)| \leq C \frac{|\bar{\rho}_j|}{\rho_0} h_j, \quad \forall x \in I_j.$$

Note also that for all  $x \in I_j$

$$(4.5) \quad \underline{a}_j = \min\{a(x) \mid x \in I_j\} = a_{j-1/2}(1+O(h_j^2)), \text{ as } h_j \rightarrow 0.$$

Since for any  $v \in E_0(I_j)$

$$I_j \|v\|_E^2 = \left[ \int_{I_j} a(x) v'(x)^2 dx \right] (1+O(h_j^2)), \text{ as } h_j \rightarrow 0,$$

it follows from (4.3) and (4.5) that

$$\begin{aligned} I_j \|\phi_{1,j}\|_E^2 &= \left[ \sup_{v \in E_0(I_j)} \frac{1}{I_j \|v\|_E} |F_{\bar{\rho}_j}(v)| \right]^2 \\ &= \frac{\bar{\rho}_j^2}{\underline{a}_j} \left[ \sup_{v \in E_0(I_j)} \left\{ \frac{|\int_{I_j} v dx|^2}{\int_{I_j} (v')^2 dx} \right\} \right] (1+O(h_j^2)) \\ &= \frac{\bar{\rho}_j^2}{\underline{a}_j} \left[ \int_{I_j} (\tilde{v}')^2 dx \right] (1+O(h_j^2)) \end{aligned}$$

where  $\tilde{v} \in E_0(I_j)$  is the solution of

$$-\tilde{v}'' = 1, \quad \tilde{v}(x_{j-1}) = \tilde{v}(x_j) = 0.$$

This implies that

$$(4.6) \quad I_j \|\phi_{1,j}\|_E^2 = \frac{1}{12} \frac{\bar{\rho}_j^2}{\underline{a}_{j-1/2}} h_j^3 (1+O(h_j^2)), \text{ as } h_j \rightarrow 0, \quad j = 1, \dots, m.$$

In order to estimate  $\phi_{2,j}$  we proceed as in the proof of (3.14). The smallest eigenvalue of the operator  $L$  on  $I_j$  is bounded below by



$a_j \pi^2 / h_j^2$ . Hence by (4.3) it follows that

$$(4.7) \quad I_j \|\phi_{2,j}\|_E^2 = F_{\sigma_j}(\phi_{2,j}) \leq I_j \|\phi_{2,j}\|_0 I_j \|\sigma_j\|_0 \leq \frac{h_j^2}{a_j \pi^2} I_j \|\sigma_j\|_0^2,$$

which together with (4.4) implies that

$$(4.8) \quad I_j \|\phi_{2,j}\|_E^2 \leq \frac{C}{\rho_0^2 \pi^2} \frac{\bar{\rho}_j^2}{a_{j-1/2}} h_j^5 (1+O(h_j)), \text{ as } h_j \rightarrow 0, j = 1, \dots, m.$$

Combining this with (4.6) we obtain--with some  $|\alpha| \leq 1$ --

$$(4.9) \quad \begin{aligned} I_j \|\phi_j\|_E^2 &= I_j \|\phi_{1,j}\|_E^2 + 2\alpha I_j \|\phi_{1,j}\|_E I_j \|\phi_{2,j}\|_E + I_j \|\phi_{2,j}\|_E^2 \\ &= \frac{1}{12} \frac{\bar{\rho}_j^2}{a_{j-1/2}} h_j^3 (1+O(h_j)), \text{ as } h_j \rightarrow 0, j = 1, \dots, m. \end{aligned}$$

By definition (3.17) of  $Q(\Delta)$  this proves the lemma.

A partition  $\Delta$  shall be a  $(\xi, m)$ -partition if

$$(4.10) \quad \xi(x_j^\Delta) = \frac{j}{m}, j = 0, 1, \dots, m$$

for some function

$$(4.11) \quad \xi \in P^2(I, \Delta), \xi'(x) \geq \delta > 0, \forall x \in I_j, j = 1, \dots, m, \xi(0) = 0, \xi(1) = 1.$$

Note that any partition  $\Delta$  is a  $(\xi, m)$ -partition for the piecewise linear function  $\xi \in S(I, \Delta)$  defined by

$$(4.12) \quad \xi(x) = \frac{j-1}{m} + \frac{1}{mh_j} (x - x_{j-1}), x \in I_j, j = 1, \dots, m.$$

For a  $(\xi, m)$ -partition we have

$$(4.13) \quad \frac{1}{m} = \int_{I_j} \xi'(t) dt = h_j \xi'(x_{j-1/2})(1+O(h_j)), \text{ as } h_j \rightarrow 0$$

where the constant in the bound of the 0-term depends on  $\xi$  but not on  $m$ .

In terms of  $(\xi, m)$ -partitions our Lemma 4.1 can be rewritten as follows:

Theorem 4.2: For the  $(\xi, m)$ -partition  $\Delta$  we have

$$(4.14) \quad Q(\Delta)^2 = \frac{1}{12m^2} \left[ \int_0^1 \left( \frac{\rho(x)}{\xi'(x)} \right)^2 \frac{1}{a(x)} dx \right] (1+O(\bar{h}(\Delta))), \text{ as } \bar{h}(\Delta) \rightarrow 0$$

where the constant in the bound of the 0-term depends on  $a$ ,  $b$ ,  $f$  and  $\xi$  but not on  $m$ .

Because  $\rho \in C^1(I)$  and  $1/\xi' \in P^1(I, \Delta)$ , the proof follows directly from the fact that the expression for  $Q(\Delta)^2$  in Lemma 4.1 is a Riemann sum of (4.14).

By combining Theorems 3.1, 3.2, and 4.2 we obtain the following result:

Theorem 4.3: For the  $(\xi, m)$ -partition  $\Delta$  the error satisfies

$$(4.15) \quad \|e\|_E^2 = \frac{1}{12m^2} \left[ \int_0^1 \left( \frac{\rho(x)}{\xi'(x)} \right)^2 \frac{1}{a(x)} dx \right] (1+O(\bar{h}(\Delta))), \text{ as } \bar{h}(\Delta) \rightarrow 0$$

where the constant in the bound of the 0-term depends on  $a$ ,  $b$ ,  $f$  and  $\xi$  but not on  $m$ .

#### 4.2 Optimal Partitions

The error formula of Theorem 4.3 suggests that we consider minimizing the variational integral

$$(4.16a) \quad J(\xi) = \int_0^1 \left( \frac{\rho(x)}{\xi'(x)} \right)^2 \frac{1}{a(x)} dx$$

subject to the boundary conditions

$$(4.16b) \quad \xi(0) = 0, \xi(1) = 1.$$

The Euler equation is directly solvable and the functions

$$(4.17) \quad \xi(x, \gamma) = \gamma \int_0^x \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt, \quad x \in I,$$

form a field of extremals in  $0 < x, \gamma < 1$ . A standard application of the E-function test (see, e.g., [1]) then proves the following result:

Theorem 4.4: For all functions (4.10) we have

$$(4.18) \quad J(\xi) \geq J(\xi_0) = 1/\gamma_0^3$$

where  $\xi_0 = \xi(., \gamma_0)$  with

$$(4.19) \quad \gamma_0 = \left[ \int_0^1 \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt \right]^{-1}.$$

Moreover, equality holds in (4.18) exactly when  $\xi = \xi_0$ .

Note that the function  $\xi_0$  belongs to the class of functions (4.11); in fact, we have  $\xi_0 \in C^2(I)$  and  $\xi_0'(x) \geq \delta > 0$  for  $x \in I_j$  because  $|\rho(x)| \geq \rho_0$  in that interval. For the partition  $\Delta_0$  given by  $\xi_0$  we obtain from Theorems 4.3 and 4.4 the following error formula.

Theorem 4.5: The  $(\xi_0, m)$ -partition  $\Delta_0$  is asymptotically optimal with

$$(4.20) \quad \|e\|_E^2 = \frac{1}{12m^2 \gamma_0^3} (1 + O(\bar{h}(\Delta_0))), \text{ as } \bar{h}(\Delta) \rightarrow 0$$

where the constant in the bound of the  $O$ -term depends on  $\xi_0$  but not on  $m$ .

By "asymptotic optimality" we mean here that for any other partition

with sufficiently small  $\bar{h}$  the error is larger than (4.20).

For any  $(\xi, m)$ -partition  $\Delta$  set

$$(4.21) \quad \vartheta_j(\xi, m)^2 = \frac{1}{12m^2} \int_{I_j} \left[ \frac{\rho(t)}{\xi'(t)} \right]^2 \frac{1}{a(t)} dt, \quad j = 1, \dots, m.$$

These  $\vartheta_j$  are related to the functions  $\phi_j$  of (3.13) by

$$(4.22) \quad I_j \|\phi_j\|_E^2 = \vartheta_j(\xi, m)^2 (1 + O(h_j)), \quad \text{as } h_j \rightarrow 0, \quad j = 1, \dots, m.$$

This follows directly from the fact that the expression (4.9) for the norms of  $\phi_j$  is--up to a factor  $(1 + O(h_j))$ --a Riemann sum of (4.22).

For the optimal partition  $\Delta_0$  we have

$$(4.23) \quad \begin{aligned} \vartheta_j(\xi_0, m)^2 &= \frac{1}{12m^2 \gamma_0^2} \int_{I_j} \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt \\ &= \frac{1}{12m^2 \gamma_0^3} \int_{I_j} \xi_0'(t) dt = \frac{1}{12m^2 \gamma_0^3}, \quad j = 1, \dots, m. \end{aligned}$$

In other words, for  $\xi_0$  all  $\vartheta_j$  are exactly equal and--by (4.22)--all  $I_j \|\phi_j\|_E^2$  are asymptotically equal.

Since the  $\vartheta_j$  are not readily computable, we turn now to the quantities

$$(4.24) \quad \mu_j(\xi, m) = I_j \|z_j\|_E, \quad j = 1, \dots, m$$

which can be calculated. For the optimal partition  $\Delta_0$  we obtain from (3.16), (4.13) and Theorem 4.5 that

$$(4.25) \quad I_j \|\psi_j\|_E^2 = \frac{1}{12m^3 \gamma_0^3} (1+O(\bar{h}(\Delta_0))) (O(\bar{h}(\Delta_0))), \text{ as } \bar{h}(\Delta_0) \rightarrow 0$$

where the constant in the 0-term depends on  $a, b, f$  and  $\xi_0$  but not on  $m$ .

Thus it follows from (4.22), (4.23) and (4.25) that

$$\begin{aligned} \mu_j(\xi_0, m)^2 &= I_j \|\varphi_j + \psi_j\|_E^2 = \frac{1}{12m^3 \gamma_0^3} (1+O(\bar{h})) \\ &\quad + \frac{1}{\sqrt{12} m^{3/2} \gamma_0^{3/2}} \frac{1}{\sqrt{12} m^{3/2} \gamma_0^{3/2}} (1+O(\bar{h})) O(\bar{h}^{1/2}) \\ &\quad + \frac{1}{12m^3 \gamma_0^3} (1+O(\bar{h})) O(\bar{h}) = \frac{1}{12m^3 \gamma_0^3} (1+O(\bar{h}^{1/2})) . \end{aligned}$$

We summarize this in the following form:

Theorem 4.6: For the optimal partition  $\Delta_0$  we have

$$(4.26) \quad \mu_j(\xi_0, m)^2 = \frac{1}{12m^3 \gamma_0^3} (1+O(\bar{h}(\Delta_0)^{1/2})), \text{ as } \bar{h}(\Delta_0) \rightarrow 0 ,$$

where the constant in the bound of the 0-term depends on  $\xi_0$  but not on  $m$ .

Thus, we see that also the quantities  $\mu_j(\xi_0, m)$  are asymptotically equal.

From (4.13) we find for the steps  $h_j$  of the optimal partition  $\Delta_0$  that

$$(4.27) \quad h_j = \frac{1}{\gamma_0^m} \left[ \frac{a_j - 1/2}{\bar{\rho}_j^2} \right]^{1/3} (1+O(\frac{1}{m})), \text{ as } m \rightarrow \infty$$

and hence that

$$\frac{h(\Delta_0)}{\bar{h}(\Delta_0)} \geq \lambda_0 > 0$$

with a constant  $\lambda_0$  that does not depend on  $m$  but only on the problem and  $\rho_0$ . In other words, the optimal partition is  $(\lambda_0, 1)$ -regular.

Conversely, it turns out that asymptotically the optimal partition  $\Delta_0$  is characterized by the asymptotic equality of the  $\mu_j$ . This is the content of the following theorem.

Theorem 4.7: For the partition  $\Delta$  suppose that

$$(4.28) \quad I_j \|z_j(\Delta)\|_E = \mu(1+O(\bar{h}^{1/2})), \text{ as } \bar{h}(\Delta) \rightarrow 0, j = 1, \dots, m,$$

where  $\mu$  does not depend on  $j$ . Then

$$(4.29) \quad I \|e(\Delta)\|_E = I \|e(\Delta_0)\|_E(1+O(\bar{h}^{1/2})), \text{ as } \bar{h}(\Delta) \rightarrow 0, j = 1, \dots, m,$$

and

$$(4.30) \quad |x_j^{\Delta} - x_j^{\Delta_0}| = O(\bar{h}^{1/2}), \text{ as } \bar{h}(\Delta) \rightarrow 0, j = 1, \dots, m.$$

Proof: We show first that  $\Delta$  is  $(\lambda, 1)$ -regular. For ease of notation let

$$\eta_j = \left[ \frac{1}{12} \frac{\bar{\rho}_j^2}{a_{j-1/2}} \right]^{1/2}, j = 1, \dots, m.$$

For any  $j = 1, \dots, m$ , we have by (4.9), with some  $|\alpha| \leq 1$ ,

$$\begin{aligned} (4.31) \quad I_j \|z_j\|_E^2 &= I_j \|\varphi_j + \psi_j\|_E^2 = \\ &= \eta_j^2 h_j^3 (1+O(\bar{h})) + 2\alpha \eta_j h_j^{3/2} I_j \|\psi_j\|_E (1+O(\bar{h})) + I_j \|\psi_j\|_E^2. \end{aligned}$$

Let now  $h_{j_0} = \bar{h}$ . Then (3.16) and (3.6) show that

$$I_{j_0} \|\psi_{j_0}\|_E \leq C \bar{h} I_{j_0} \|e\|_E \leq C \bar{h}^2$$

or

$$(4.32) \quad I_{j_0} \|\psi_{j_0}\|_E = \bar{h}^{3/2} O(\bar{h}^{1/2}), \text{ as } \bar{h} \rightarrow 0.$$

Hence we obtain from (4.31) and (4.28) that

$$\mu^2(1+O(\bar{h}^{1/2})) = \eta_{j_0}^2 \bar{h}^3(1+O(\bar{h})) + \bar{h}^3 O(\bar{h}^{1/2}) + \bar{h}^3 O(\bar{h})$$

or

$$(4.33) \quad \mu^2 = \eta_{j_0}^2 \bar{h}^3(1+O(\bar{h}^{1/2})), \text{ as } \bar{h} \rightarrow 0.$$

Now let  $h_{j_1} = \underline{h}$ . Then we have instead of (4.32)

$$I_{j_1} \|\psi_{j_1}\|_E = \underline{h} \bar{h}^{1/2} O(\bar{h}^{1/2}),$$

and hence by (4.31), (4.28), and (4.33)

$$(4.34) \quad \eta_{j_0}^2 \bar{h}^3(1+O(\bar{h}^{1/2})) = \eta_{j_1}^2 \underline{h}^3(1+O(\bar{h})) + 2\alpha \eta_{j_1} \underline{h}^{5/2} \bar{h}^{1/2} O(\bar{h}^{1/2}) + \underline{h}^2 \bar{h} O(\bar{h}).$$

In other words,  $z^2 = \bar{h}/\underline{h}$  satisfies the polynomial equation

$$(4.35) \quad \eta_{j_0}^2 (1+O(\bar{h}^{1/2})) z^6 + O(\bar{h}) z^2 + O(\bar{h}^{1/2}) z - \eta_{j_1}^2 (1+O(\bar{h})) = 0.$$

By comparing (4.35) with the equation

$$(4.36) \quad \eta_{j_0}^2 (1+O(\bar{h}^{1/2})) z^6 - \eta_{j_1}^2 (1+O(\bar{h})) = 0,$$

Rouché's theorem shows that for sufficiently small  $\bar{h}$  we have

$$(4.37) \quad |z| \leq 2 \left( \frac{\eta_{j1}}{\eta_{j0}} \right)^{1/3}.$$

This shows that  $\Delta$  is indeed  $(\lambda, 1)$ -regular.

Our assumptions about  $a$  and  $\rho$  ensure that for any fixed  $j$ ,  $1 \leq j \leq m$ ,

$$\sum_{i=1}^m \eta_i^2 h_i \leq C \eta_j^2.$$

We represent  $\Delta$  by the piecewise linear function  $\xi$  of (4.12). Then from Theorems 3.1, 3.2, and Lemma 4.1 it follows that

$$(4.38) \quad I \|e\|_E^2 = \left[ \sum_{i=1}^m \eta_i^2 h_i^3 \right] (1+O(\bar{h})) \leq C \eta_j^2 \bar{h}^2,$$

and hence, by (3.16) that

$$(4.39) \quad I_j \|\psi_j\|_E^2 \leq C h_j^2 I \|e\|_E^2 \leq C \eta_j^2 h_j^2 \bar{h}^2 \leq C \eta_j^2 h_j^3 \bar{h},$$

where in the last inequality the regularity of  $\Delta$  was used. Therefore, we have

$$I_j \|\psi_j\|_E = \eta_j h_j^{3/2} O(\bar{h}^{1/2})$$

whence by (4.31)

$$I_j \|z_j\|_E^2 = \eta_j^2 h_j^3 [(1+O(\bar{h})) + O(\bar{h}^{1/2}) + O(\bar{h})] = \eta_j^2 h_j^3 (1+O(\bar{h}^{1/2})),$$

that is, by (4.28)

$$(4.40) \quad \mu^2 = \eta_j^2 h_j^3 (1+O(\bar{h}^{1/2})).$$



Since

$$\left[ 12\eta_j^2 \right]^{1/3} h_j = \left[ \frac{\bar{\rho}_j^2}{a_{j-1/2}} \right]^{1/3} h_j = \left[ \int_{I_j} \left( \frac{\rho(t)^2}{a(t)} \right)^{1/3} dt \right] (1+O(\bar{h})), \text{ as } \bar{h} \rightarrow 0 ,$$

(4.40) implies that

$$(4.41) \quad \int_{I_j} \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt = (12\mu^2)^{1/3} (1+O(\bar{h}^{-1/2}))$$

which by (4.19) shows that

$$(4.42) \quad \frac{1}{r_0} = \int_0^1 \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt = m(12\mu^2)^{1/3} (1+O(\bar{h}^{-1/2})) .$$

By (4.38) and (4.40) we obtain now

$$I \|e\|_E^2 = m^2 (1+O(\bar{h}^{-1/2})) ,$$

which by (4.42) and Theorem 4.5 gives (4.29). Finally from (4.41) it follows by summation over the first  $j$  intervals

$$\int_0^{x_j^\Delta} \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt = \frac{j}{mr_0} (1+O(\bar{h}^{-1/2})) .$$

On the other hand, we obtain from (4.23)

$$\int_0^{x_j^\Delta} \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt = \frac{j}{mr_0} (1+O(\bar{h}^{-1/2})) .$$

Thus we have

$$\left| \int_{x_j^{\Delta_0}}^{x_j^\Delta} \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt \right| = O(\bar{h}^{-1/2}) ,$$

which implies (4.30).

### 4.3 Computational Aspects

Suppose that the optimal partition function  $\xi_0$  is changed to some partition function  $\xi = \xi_0 + \varepsilon$  satisfying (4.11). By (4.13a) the derivatives control the stepsizes, and hence we assume that  ${}_I\|\varepsilon'\|_C$  is small. For any given  $m$ , let  ${}_I\|e\|_E$  and  ${}_I\|e_0\|_E$  be the errors associated with the  $(\xi, m)$ -partition and  $(\xi_0, m)$ -partition, respectively. Then by Theorem 4.3 and (4.16a) we have

$$|{}_I\|e\|_E^2 - {}_I\|e_0\|_E^2| = O\left(\frac{1}{m^2}\right)|J(\xi) - J(\xi_0)|, \text{ as } m \rightarrow \infty.$$

Since the variational integral  $J$  is stationary at  $\xi_0$  we have  $J'(\xi_0) = 0$  and hence it follows from the mean-value theorem that

$$|J(\xi) - J(\xi_0)| = O({}_I\|\varepsilon'\|_C^2).$$

Therefore, because (4.15) implies that

$${}_I\|e\|_E + {}_I\|e_0\|_E = O\left(\frac{1}{m}\right), \text{ as } m \rightarrow \infty,$$

we obtain

$$|{}_I\|e\|_E - {}_I\|e_0\|_E| = O\left(\frac{1}{m}\right)O({}_I\|\varepsilon'\|_C^2) \text{ as } {}_I\|\varepsilon'\|_C \rightarrow 0, m \rightarrow \infty.$$

In other words, a change of  $\xi_0'$  by some small  ${}_I\|\varepsilon'\|_C$  leads to a change of the error proportional to  ${}_I\|\varepsilon'\|_C^2$ . This shows that the value of the optimal error is rather stable under perturbations of the optimal partition. On the other hand, the optimal partition itself is not too stable and hence needs to be computed only with relatively low accuracy.

By Theorem 4.7 the optimal partition is characterized by the asymptotic equality of the quantities  $\mu_j = \mu_j(\xi_0, m)$  of (4.24). Let  $r_j$  again denote

the residual  $r = L(u_\Delta) - f$  on the subinterval  $I_j$  and set

$$(4.43) \quad v_j^2 = \int_{I_j} r_j(x)^2 dx, \quad j = 1, \dots, m.$$

We call the quantities

$$(4.44) \quad \epsilon_j^2 = \frac{1}{12} \frac{v_j^2 h_j^2}{a_{j-1/2}}, \quad j = 1, 2, \dots, m$$

the error indicators for the intervals  $I_1, \dots, I_m$ , and set

$$(4.45) \quad \epsilon(\Delta) = \left( \sum_{j=1}^m \epsilon_j^2 \right)^{1/2}.$$

All these quantities are directly computable once the finite element solution is known. The next theorem shows that  $\epsilon(\Delta)$  is asymptotically equal to the error  $\|e\|_E$ .

Theorem 4.8: a) For any partition  $\Delta$  we have

$$(4.46a) \quad \|e\|_E^2 = \epsilon(\Delta)^2 (1 + O(\bar{h})), \quad \text{as } \bar{h}(\Delta) \rightarrow 0.$$

b) If  $\Delta$  is  $(\lambda, \kappa)$ -regular,  $1 \leq \kappa < 2$ , then

$$(4.46b) \quad v_j^2 = \rho_j^2 h_j^2 (1 + O(\bar{h}^\epsilon)), \quad \text{as } \bar{h}(\Delta) \rightarrow 0, \quad j = 1, \dots, m,$$

where  $\epsilon = 1 - \kappa/2$ .

Proof: a) The definition of  $\bar{\rho}_j$  given in Lemma 4.1 provides that

$$|\bar{\rho}_j| = \max_{x \in I_j} |\rho_j(x)| \geq \rho_0 > 0,$$

and by (4.4), (3.15), and (3.6) we have

$$|\sigma_j(x)| \leq C \frac{1}{\rho_0} \bar{\rho}_j h_j, \quad \| \tau_j \|_0 \leq C \|e\|_E.$$

Hence by (3.11) and (4.2) we obtain with some  $|a_i| \leq 1$ ,  $i = 1, 2, 3$ , that

$$\begin{aligned}
 v_j^2 &= \int_{I_j} (\bar{\rho}_j + \sigma_j(x) + \tau_j(x))^2 dx \\
 (4.47) \quad &= \frac{\bar{\rho}_j^2 h_j}{\rho_0} + \frac{1}{2} \frac{\bar{\rho}_j^2 h_j^2}{\rho_0} O(h_j) + I_j \|\tau_j\|_0^2 \\
 &\quad + 2a_1 \frac{1}{\rho_0} \bar{\rho}_j^2 h_j O(h_j) + 2a_2 \bar{\rho}_j h_j^{1/2} I_j \|\tau_j\|_0 + 2a_3 \frac{1}{\rho_0} \bar{\rho}_j h_j^{3/2} I_j \|\tau_j\|_0
 \end{aligned}$$

With

$$S(\Delta)^2 = \frac{1}{12} \sum_{j=1}^m \frac{\bar{\rho}_j^3 h_j^3}{a_{j-1/2}}, \quad T(\Delta)^2 = \sum_{j=1}^m I_j \|\tau_j\|_0^2$$

it then follows that

$$\begin{aligned}
 \epsilon(\Delta)^2 &= S(\Delta)^2 + \frac{1}{\rho_0^2} S(\Delta)^2 O(\bar{h}^2) + T(\Delta)^2 O(\bar{h}^2) \\
 &\quad + 2 \frac{a_1}{\rho_0} S(\Delta)^2 O(\bar{h}) + 2a_2 S(\Delta) T(\Delta) O(\bar{h}) + 2a_3 S(\Delta) T(\Delta) O(\bar{h}^2).
 \end{aligned}$$

This proves (4.46a) since by Lemma 4.1 and Theorems 3.1 and 3.2

$$S(\Delta)^2 = I \|\mathbf{e}\|_E^2 (1 + O(\bar{h})), \quad T(\Delta)^2 \leq C \sum_{j=1}^m I_j \|\mathbf{e}\|_E^2 = C I \|\mathbf{e}\|_E^2.$$

b) For  $(\lambda, \kappa)$ -regular  $\Delta$  with  $1 \leq \kappa < 2$  we have by (3.6)

$$I_j \|\tau_j\|_0 \leq C I_j \|\mathbf{e}\|_E \leq C \bar{h} \leq C h_j^{1/2} \bar{h}^\epsilon.$$

Hence (4.47) implies that

$$v_j^2 = \frac{\bar{\rho}_j^2 h_j}{\rho_0} \left[ 1 + \frac{1}{2} O(\bar{h}^2) + \frac{1}{2} O(\bar{h}^{2\epsilon}) + \frac{1}{\rho_0} O(\bar{h}) + \frac{1}{\rho_0} O(\bar{h}^\epsilon) + \frac{1}{\rho_0^2} O(\bar{h}^{3/2}) \right]$$

which is (4.46b).

It may be noted that in [3] we proved an upper bound for  $I\|e\|_E$  of the form (4.46b) with  $1/12$  replaced by the larger factor  $1/\pi^2$ .

Theorem 4.8 states that

$$(4.48) \quad \theta(\Delta) = \frac{I\|e\|_E}{\varepsilon(\Delta)} = 1 + O(\bar{h}(\Delta)), \quad \text{as } \bar{h}(\Delta) \rightarrow 0.$$

In other words, the effectivity quotient  $\theta$  tends to one with  $\bar{h} \rightarrow 0$ .

In contrast, the corresponding estimates in [3] only provided for  $\theta^2 \leq 12/\pi^2$ .

We expect the error indicators to be asymptotically equal to the quantities  $\mu_j$  of (4.24). Theorem 4.9 below shows that this is indeed correct for regular partitions. Hence our aim is to construct such partitions for which all  $\varepsilon_j$  are asymptotically equal. It turns out that --as before-- these partitions are automatically regular.

Theorem 4.9: (a) Let  $\Delta$  be a  $(\lambda, \kappa)$ -regular partition with  $1 \leq \kappa < 2$ .

Then

$$(4.49) \quad I_j \|z_j\|_E^2 = \varepsilon_j^2 (1 + O(\bar{h}^\varepsilon)), \quad \text{as } \bar{h}(\Delta) \rightarrow 0,$$

with  $\varepsilon = 1 - \kappa/2$ .

(b) All partitions  $\Delta$  for which

$$(4.50) \quad \varepsilon_j = \mu(1 + o(1)) \quad \text{as } \bar{h}(\Delta) \rightarrow 0, \quad j = 1, \dots, m,$$

with  $\mu$  independent of  $j$ , are  $(\lambda, 1)$ -regular.

Proof: As in the case of (4.39) it follows from (3.16) and the assumed regularity of  $\Delta$  that

$$I_j \|\psi_j\|_E^2 \leq C \eta_j^2 h_j^{2-2\varepsilon} \leq C \eta_j^2 h_j^{3-2\varepsilon},$$

and thus

$$I_j \|\psi_j\|_E = \eta_j h_j^{3/2} O(\bar{h}^\epsilon), \text{ as } \bar{h}(\Delta) \rightarrow 0.$$

Now (4.31) shows that

$$\begin{aligned} I_j \|z_j\|_E^2 &= \eta_j^2 h_j^3 (1+O(\bar{h})) + 2\alpha \eta_j^2 h_j^3 O(\bar{h}^\epsilon) (1+O(\bar{h})) + \eta_j^2 h_j^3 O(\bar{h}^{2\epsilon}) \\ &= \eta_j^2 h_j^3 (1+O(\bar{h}^\epsilon)). \end{aligned}$$

Since by Theorem 4.8(b)

$$(4.51) \quad \eta_j^2 h_j^3 = \frac{1}{12} \frac{\rho_j^2 h_j^3}{a_{j-1/2}} = \epsilon_j^2 (1+O(\bar{h}^\epsilon)),$$

this proves (4.49).

(b) Because, generally,

$$I_j \|\tau_j\|_E \leq C I_j \|e\|_E \leq C \bar{h},$$

it follows from (4.47) that

$$v_j^2 = \rho_j^2 h_j^3 (1+O(h_j)) + O(\bar{h}^{3/2})$$

whence

$$\mu^2(1+o(1)) = \epsilon_j^2 = \eta_j^2 h_j^3 + h_j^2 O(\bar{h}^{1/2}).$$

Now suppose that (4.50) holds for  $\Delta$  and that  $h_{j_0} = \bar{h}$ ; then

$$\mu^2(1+o(1)) = \bar{h}^3 (\eta_{j_0}^2 + O(\bar{h}^{1/2})).$$

Similarly, for  $h_{j_1} = \underline{h}$  we obtain

$$\mu^2(1+o(1)) = \eta_{j_1}^2 \underline{h}^3 + \underline{h}^2 O(\bar{h}^{1/2})$$

and hence  $z = \bar{h}/h$  satisfies

$$[\eta_{j_0}^2 + O(\bar{h}^{1/2})]z^3 = \eta_{j_1}^2 + O(\bar{h}^{1/2})z.$$

By comparing this with the polynomial for  $z$  in which the last term on the right has been dropped, it follows by Rouché's theorem that

$$|z| \leq 2 \left( \frac{\eta_{j_1}}{\eta_{j_0}} \right)^{2/3}$$

for sufficiently small  $\bar{h}$  and therefore that  $\Delta$  is  $(\lambda, 1)$ -regular.

Theorems 4.9 and 4.7 confirm that, as expected, our aim should be to construct partitions for which all  $\epsilon_j$  are asymptotically equal. Then the error of the partition will be close to the asymptotically optimal error (4.20). A natural approach for this construction is the use of an adaptive mesh refinement algorithm of the form discussed, for instance, in [4]. We shall not repeat the details.

## 5. Optimal Partitions - Part II

In the previous section we assumed that  $u_0''(x) \neq 0$  for  $x \in I$ . Clearly, this represents a severe restriction. Actually, the results are largely valid also when  $u_0''$  has zeroes in  $I$ , but the proofs become more delicate. We illustrate the approach for the frequent case when  $u_0'' \in C^1(I)$  has finitely many simple roots in  $I$ , say,

$$(5.1) \quad u_0''(\xi_k) = 0, u_0'''(\xi_k) \neq 0, k = 1, \dots, q, 0 \leq \xi_1 < \xi_2 < \dots < \xi_q \leq 1.$$

Lemma 5.1: Under the stated assumptions we have for any  $(\lambda, \kappa)$ -regular partition  $\Delta$  with  $1 \leq \kappa < 2$ ,

$$(5.2) \quad Q(\Delta)^2 = \frac{1}{12} \left[ \sum_{j=1}^m \frac{\bar{\rho}_j^2}{a_{j-1/2}} h_j^3 \right] (1 + O(\bar{h}(\Delta)^\epsilon)), \text{ as } \bar{h}(\Delta) \rightarrow 0$$

where  $\epsilon = 1 - \kappa/2$  and the constant in the bound of the 0-term depends on  $a, b$  and  $f$ .

Proof: Because of (5.1), we may choose  $c_2 \geq c_1 > 0$ ,  $\delta_0 > 0$ , such that

$$(5.3) \quad c_2 |x - \xi_k| \geq |\rho(x)| \geq c_1 |x - \xi_k|, \forall |x - \xi_k| \leq \delta_0, k = 1, \dots, q.$$

For any  $\delta > 0$  we introduce the sets

$$(5.4) \quad I_\delta = \{x \in \bar{I} \mid |x - \xi_k| < \delta \text{ for some } \xi_k\}, \quad I_\delta^c = \bar{I} \setminus I_\delta,$$

$$J_\delta = \{j=1, \dots, m; I_j \cap I_\delta \neq \emptyset\}, \quad J_\delta^c = \{1, \dots, m\} \setminus J_\delta.$$



We assume that  $\delta_0 \leq (8q)^{-1}$  and hence that

$$\sum_{j \in J_{\delta_0}} h_j \leq 2(\delta_0 + \bar{h})q \leq 4\delta_0 q \leq \frac{1}{2}, \text{ for } \bar{h} \leq \delta_0.$$

Since

$$\min \{ |\rho(x)|, x \in I_{\delta_0}^c \} = \rho_0 > 0,$$

the estimate (4.9) of the proof of Lemma 4.1 holds for the subintervals  $I_j$  with  $j \in J_{\delta_0}^c$ ; that is

$$I_j \|\varphi_j\|_E^2 = \eta_j^2 h_j^3 (1+O(\bar{h})) \text{ as } \bar{h}(\Delta) \rightarrow 0, \quad j \in J_{\delta_0}^c.$$

Hence for  $\bar{h} \leq \delta_0$  we have

$$(5.5) \quad Q(\Delta)^2 \geq \left[ \sum_{j \in J_{\delta_0}^c} \eta_j^2 h_j^3 \right] (1+O(\bar{h})) \geq C \rho_0^2 \lambda^2 \bar{h}^{-2\kappa} \sum_{j \in J_{\delta_0}^c} h_j \geq C \bar{h}^{-2\kappa}.$$

Now consider the sets (5.4) with  $\delta = \bar{h}^{\kappa/2} \leq \delta_0$ . Then (5.3) implies that

$$|\bar{\rho}_j| \geq c_1 \bar{h}^{\kappa/2}, \quad j \in J_{\delta}^c,$$

and hence (4.8) modifies to

$$(5.6) \quad I_j \|\varphi_{2,j}\|_E^2 \leq C \frac{1}{\bar{h}^{\kappa}} \eta_j^2 h_j^5 (1+O(\bar{h})) \leq C \eta_j^2 h_j^3 \bar{h}^{-2\epsilon}, \text{ as } \bar{h}(\Delta) \rightarrow 0, \quad j \in J_{\delta}^c.$$

On the other hand we have

$$|\sigma_j(x)| \leq 2 \max_{x \in I_j} |\rho(x)| \leq C \bar{h}^{\kappa/2}, \quad j \in J_{\delta}$$

whence by (4.7) and (5.5)

$$(5.7) \quad I_j \|\phi_{2,j}\|_E^2 \leq Ch_j^3 \bar{h}^K (1+O(\bar{h})) \leq CQ(\Delta)^2 \bar{h}^{2\epsilon} h_j, \text{ as } \bar{h}(\Delta) \rightarrow 0, j \in J_\delta.$$

For ease of notation wet

$$S(\Delta)^2 = \sum_{j=1}^m \eta_j^2 h_j^3, \quad R(\Delta)^2 = \sum_{j=1}^m I_j \|\phi_{2,j}\|_E^2.$$

Then it follows from (5.6) and (5.7) that

$$(5.8) \quad \begin{aligned} R(\Delta)^2 &\leq C \left[ \sum_{j \in J_\delta^c} \eta_j^2 h_j^3 \right] \bar{h}^{2\epsilon} + CQ(\Delta)^2 \bar{h}^{2\epsilon} \sum_{j \in J_\delta} h_j \\ &\leq C[S(\Delta)^2 + Q(\Delta)^2] \bar{h}^{2\epsilon}. \end{aligned}$$

In the case of Lemma 4.1 the assumption  $u_0'' = 0$  does not enter into the proof of (4.6), and thus we have also in the present case

$$(5.9) \quad I_j \|\phi_{1,j}\|_E^2 = \eta_j^2 h_j^3 (1+O(\bar{h})), \text{ as } \bar{h}(\Delta) \rightarrow 0, j = 1, \dots, m$$

and therefore

$$(5.10) \quad \sum_{j=1}^m I_j \|\phi_{1,j}\|_E^2 = S(\Delta)^2 (1+O(\bar{h})), \text{ as } \bar{h} \rightarrow 0.$$

Altogether, with some suitable constants  $\alpha, \beta \in [-1, 1]$  it follows from (5.8) and (5.10) that

$$\begin{aligned} Q(\Delta)^2 &= \sum_{j=1}^m I_j \|\phi_{1,j}\|_E^2 + 2\alpha \left( \sum_{j=1}^m I_j \|\phi_{1,j}\|_E^2 \right)^{1/2} \left( \sum_{j=1}^m I_j \|\phi_{2,j}\|_E^2 \right)^{1/2} + R(\Delta)^2 \\ &= S(\Delta)^2 (1+O(\bar{h})) + \alpha C S(\Delta) (1+O(\bar{h})) (S(\Delta)^2 + Q(\Delta)^2)^{1/2} \bar{h}^\epsilon \\ &\quad + \beta C (S(\Delta)^2 + Q(\Delta)^2) \bar{h}^{2\epsilon}. \end{aligned}$$

After separating the middle term on the right and squaring, we obtain the equation

$$\begin{aligned} & CS(\Delta)^2(S(\Delta)^2+Q(\Delta)^2)\bar{h}^{2\epsilon} \\ &= Q(\Delta)^4(1+O(\bar{h}^{2\epsilon}))-2Q(\Delta)^2S(\Delta)^2(1+O(\bar{h}^{2\epsilon})) + S(\Delta)^4(1+O(\bar{h}^{2\epsilon})), \end{aligned}$$

which has the solutions

$$Q(\Delta)^2 = S(\Delta)^2(1+O(\bar{h}^{2\epsilon})) \pm [S(\Delta)^4O(\bar{h}^{2\epsilon})]^{1/2}, \text{ as } h(\Delta) \rightarrow 0,$$

and hence proves (5.2).

Now the theory of Section 4 can be carried over to the present case. As before, we consider partition functions  $\xi$ , but here we need to weaken (4.11) by requiring instead of  $\xi'(x) \geq \delta > 0$  on each  $I_j$  that  $(\rho(x)/\xi'(x))^2$  is Riemann integrable on each subinterval. Moreover, we assume that for given  $\xi$  and  $m \rightarrow \infty$  the resulting  $(\xi, m)$ -partitions are  $(\lambda, \kappa)$ -regular with  $1 \leq \kappa < 2$ .

Then as in the case of Theorem 4.2 it follows that

$$(5.11) \quad Q(\Delta)^2 = \frac{1}{12m^2} \left[ \int_0^1 \left( \frac{\rho(x)}{\xi'(x)} \right)^2 \frac{1}{a(x)} dx \right] (1+O(\bar{h}(\Delta)^\epsilon)), \text{ as } \bar{h}(\Delta) \rightarrow 0,$$

where the constant in the bound of the 0-term depends on  $a, b, f$  and  $\xi$  but not on  $m$ . This suggests again consideration of the variational problems (4.16a/b) and hence of the optimal partition function  $\xi_0$  of Theorem 4.4. Clearly

$$\frac{\rho(x)}{\xi_0'(x)} = \frac{1}{r_0} (\rho(x)a(x))^{1/3}$$

is Riemann-integrable on  $\bar{I}$ .

We show first that the  $(\xi_0, m)$ -partitions are  $(\lambda, 5/3)$ -regular. Since

$$(5.12) \quad |\rho(x)| \geq C \min_k \{1, \min |x - \xi_k|\}, \quad \forall x \in I,$$

with some  $C > 0$ , it follows that

$$C_1 h_j \geq \int_{I_j} \xi'_0(x) dx = \frac{1}{m} \geq C_2 \int_{I_j} \rho(x)^{2/3} dx \geq C_3 h_j^{5/3},$$

where again  $C_3 > 0$ . This implies that  $\bar{h} \leq C m^{-3/5}$  as well as  $1/m \leq C \bar{h}$ , and hence the partition is indeed  $(\lambda, 5/3)$ -regular. Now Theorem 4.5 is easily shown to hold with  $O(\bar{h})$  replaced by  $O(\bar{h}^{1/6})$ .

For any  $I_j$  which intersects  $I_\delta^C$  with  $\delta = \bar{h}^{5/6}$  we can use (5.6) in the estimates leading to Theorem 4.6 to obtain

$$(5.13) \quad \mu_j(\xi_0, m)^2 = \frac{1}{12m^3 \gamma_0^3} (1 + O(\bar{h}^{1/6})), \text{ as } \bar{h} \rightarrow 0, j \in J_\delta^C, \delta = \bar{h}^{5/6}.$$

In other words, for the optimal partition the  $\mu_j$  are asymptotically equal for all intervals which are not too close to a root of  $u_0''$ . As the numerical examples of Section 6 show, the  $\mu_j(\xi_0, m)$  for the intervals close to roots are generally larger and the ratio of the largest to the smallest of these values does not tend to one for  $m \rightarrow \infty$ .

The analog of Theorem 4.7 is somewhat more complicated. We formulate it as the following theorem.

Theorem 5.2: For the partition  $\Delta$  suppose that

$$(5.14) \quad \|z_j(\Delta)\|_E = \mu(1 + O(\bar{h}^\epsilon)), \quad \text{as } \bar{h}(\Delta) \rightarrow 0, j = 1, \dots, m,$$

where  $\mu$  does not depend on  $j$  and  $\epsilon = 1/12$ . Then  $\Delta$  is  $(\lambda, 5/3)$ -regular and

$$(5.15) \quad \|e(\Delta)\|_E \leq \|e(\Delta_0)\|_E (1 + O(\bar{h}^\epsilon)), \quad \text{as } \bar{h}(\Delta) \rightarrow 0.$$

Proof: 1) We show first that

$$(5.16) \quad I_j \|\phi_j\|_E \geq C h_j^{5/2}, \quad j = 1, \dots, m,$$

with some  $C > 0$  which depends only on  $\rho$ . From (3.13) it follows that

$$I_j \|\phi_j\|_E = \sup \left\{ \frac{1}{I_j \|\phi_j\|_E} \left| \int_{I_j} \rho(x) v(x) dx \right| \mid \forall v \in E_0(I_j) \right\}.$$

Clearly, for given  $\gamma, \delta$  the function

$$v_j(x) = \left( \frac{1}{2} \gamma h_j + \frac{1}{6} \delta h_j^2 \right) (x - x_{j-1}) - \frac{1}{2} \gamma (x - x_{j-1})^2 - \frac{1}{6} \delta (x - x_{j-1})^3, \quad \forall x \in I_j$$

belongs to  $E_0(I_j)$  and a short calculation shows that

$$C_1 \alpha_j^2 h_j^3 \leq \int_{I_j} v_j'(x)^2 dx \leq C_2 \alpha_j^2 h_j^3, \quad \int_{I_j} v_j(x)^2 dx \leq C_3 \alpha_j^2 h_j^5$$

where

$$\alpha_j = (\gamma^2 + (\delta h_j)^2)^{1/2}$$

and all constants are positive. Hence also

$$I_j \|\phi_j\|_E^2 \leq C \alpha_j^2 h_j^3.$$

Now with

$$\gamma = \rho(x_{j-1}), \quad \delta = \rho_j'(x_{j-1}), \quad \rho(x) = -v_j''(x) + \tilde{\rho}_j(x), \quad \forall x \in I_j$$

$$|\tilde{\rho}_j(x)| = o(h_j), \quad \text{as } h_j \rightarrow 0,$$

we obtain

$$\left| \int_{I_j} \rho(x)v(x)dx \right| \geq \int_{I_j} v'(x)^2 dx - \left( \int_{I_j} \tilde{\rho}_j(x)^2 dx \right)^{1/2} \left( \int_{I_j} v(x)^2 dx \right)^{1/2}$$

whence--for sufficiently small  $\bar{h}$ --

$$\begin{aligned} I_j \|\varphi_j\|_E &\geq C \frac{1}{\alpha_j h_j^{3/2}} \left[ \alpha_j^2 h_j^3 - o(h_j) h_j^{1/2} \alpha_j h_j^{5/2} \right] \\ (5.17) \end{aligned}$$

$$\geq C \alpha_j h_j^{3/2} \left[ 1 - \frac{1}{\alpha_j} o(h_j) \right]$$

By assumption we have

$$(5.18) \quad u_0''(x)^2 + u_0'''(x)^2 \geq C > 0, \quad \forall x \in I$$

and hence also with some  $C > 0$

$$\alpha_j^2 = \gamma^2 + (\delta h_j)^2 \geq h_j^2 (\rho(x_{j-1})^2 + \rho'(x_{j-1})^2) \geq C h_j^2 > 0$$

which together with (5.17) proves (5.16).

2) Next we show that for sufficiently small  $h_j$

$$(5.19) \quad I \|e\|_E \geq I_j \|e\|_E \geq C h_j^{5/2}, \quad C > 0.$$

Obviously, we have

$$I_j \|e\|_E^2 \geq \inf_v \int_{I_j} a(x) [u_0'(x) - v'(x)]^2 dx$$

where the infimum is taken over all linear functions on  $I_j$ . Thus  $v'(x)$  is constant and it follows that

$$I_j \|e\|_E^2 \geq C \left[ \int_{I_j} u_0'(x)^2 dx - \frac{1}{h_j} \left( \int_{I_j} u_0'(x) dx \right)^2 \right].$$

Since

$$u_0'(x) = u_0'(x_{j-1}) + u_0''(x_{j-1})(x-x_{j-1}) + \frac{1}{2} u_0'''(x_{j-1})(x-x_{j-1})^2 + o(h_j^2), \text{ as } h_j \rightarrow 0,$$

a simple calculation shows that

$$I_j \|e\|_E^2 \geq C[u_0''(x_{j-1})^2 h_j^3 + u_0'''(x_{j-1})^2 h_j^5] - o(1)h_j^5.$$

Using (5.18) we then obtain (5.19) for sufficiently small  $h_j$ .

3) Now we can show that  $\Delta$  is a  $(\lambda, 5/3)$ -regular partition. By Theorem 3.1 and (5.14) it follows that

$$(5.20) \quad I \|e\|_E^2 = m \mu^2 (1 + (\bar{h})^{1/12}),$$

and thus by (5.19) that

$$(5.21) \quad \bar{h} \leq C m^{1/5} \mu^{2/5}.$$

On the other hand, we have by (5.9), (4.7), (4.4), and (3.16)

$$(5.22) \quad \begin{aligned} \mu(1 + O(\bar{h}^{1/12})) &= I_j \|\varphi_{1,j} + \varphi_{2,j} + \psi_j\|_E \\ &\leq C[\eta_j h_j^{3/2} + h_j^{5/2} + h_j I_j \|e\|_E] \leq C[h_j^{3/2} + m^{1/2} \mu h_j]. \end{aligned}$$

Let  $h_{j_0} = \underline{h}$  and note that  $m \leq 1/\underline{h}$ , then

$$\mu \leq C(\underline{h}^{3/2} + \underline{h}^{1/2} \mu)$$

and hence for sufficiently small  $\bar{h}$

$$(5.23) \quad \underline{h} \geq C \mu^{2/3}.$$

On the other hand, (5.16) and (3.16) give

$$\begin{aligned} \mu(1+O(\bar{h}^{-1/12})) &\geq I_j \|\phi_j + \psi_j\|_E \geq I_j \|\phi_j\|_E - I_j \|\psi_j\|_E \\ &\geq C(h_j^{5/2} - h_j^{-1} \|e\|_E) \end{aligned}$$

and therefore, by (5.21) and (5.23), for  $h_{j_0} = \bar{h}$

$$(5.24) \quad \mu \geq C[\bar{h}^{5/2} - \bar{h} \mu^{1/2}] \geq C[\bar{h}^{5/2} - \bar{h} \mu^{2/3}]$$

If the term on the right is positive then, with some positive constant (independent of  $h$ ),  $z = \bar{h}^{1/2}$  must satisfy

$$z^5 - \mu^{2/3} z^2 = C\mu.$$

By Cauchy's rule this implies that

$$\bar{h}^{1/2} \leq \max((2C\mu)^{1/5}, (2\mu^{2/3})^{1/3}) \leq C\mu^{1/5}$$

since by (5.23) we have  $\mu \leq 1$  for sufficiently small  $\bar{h}$ . On the other hand, if the term on the right of (5.24) is negative, then we have immediately  $\bar{h} \leq \mu^{4/9} \leq \mu^{2/5}$ . Thus with (5.23) this gives indeed

$$\bar{h}^{5/3} \leq C \underline{h}.$$

4) For the proof of (5.15) let now  $\delta = \bar{h}^\epsilon$ . By (5.12) we have  $\eta_j \geq C\delta$  for



$j \in J_\delta^C$  and hence

$$\sum_{i=1}^m \eta_i^2 h_j \leq C \left( \frac{\eta_j}{\delta} \right)^2, \quad \forall j \in J_\delta^C.$$

Thus from Lemma 5.1 and Theorems 3.1, 3.2 it follows that

$$(5.25) \quad I_j \|e\|_E^2 \leq C \frac{1}{\delta^2} \eta_j^2 \bar{h}^2, \quad \forall j \in J_\delta^C,$$

and therefore (3.16) gives

$$(5.26) \quad I_j \|\psi_j\|_E \leq C \frac{\eta_j}{\delta} h_j \bar{h} < C \frac{\eta_j}{\delta} h_j^{3/2} \bar{h}^{2\epsilon} \leq C \eta_j h_j^{3/2} \bar{h}^\epsilon, \quad j \in J_\delta^C,$$

respectively. Note that in the second inequality the  $(\lambda, 5/3)$ -regularity of  $\Delta$  was used. Together with (5.6) and (5.9), (5.26) leads to

$$I_j \|z_j\|_E^2 = I_j \|\varphi_{1,j} + \varphi_{2,j} + \psi_j\|_E^2 = \eta_j^2 h_j^3 (1 + O(\bar{h}^\epsilon)), \quad j \in J_\delta^C,$$

and therefore by (5.14) to

$$(5.27) \quad \mu^2 = \eta_j^2 h_j^3 (1 + O(\bar{h}^\epsilon)), \quad j \in J_\delta^C.$$

Hence analogous to (4.41) we have

$$(5.28) \quad \int_{I_j} \left[ \frac{\rho(t)}{a(t)} \right]^{2/3} dt = (12\mu^2)^{1/3} (1 + O(\bar{h}^\epsilon)), \quad j \in J_\delta^C.$$

Let  $m_1$  and  $m_2$  denote the cardinalities of  $J_\delta^C$  and  $J_\delta$ , respectively, that is,  $m = m_1 + m_2$ . We want to show that

$$(5.29) \quad \frac{m_2}{m_1} = O(\delta) = O(\bar{h}^\epsilon).$$

For this note first that because of  $\eta_j \leq C\delta$  for  $j \in J_\delta$  we have by (5.22)

$$(5.30) \quad \mu \leq C(\delta h_j^{3/2} + h_j^{5/2} + h_j m^{1/2} \mu), \quad j \in J_\delta$$

and thus

$$(5.31) \quad h_j \geq C \min \left( \left( \frac{\mu}{\delta} \right)^{2/3}, m^{-1/2}, \mu^{2/5} \right), \quad j \in J_\delta.$$

Suppose first that

$$h_j \geq C \left( \frac{\mu}{\delta} \right)^{2/3}, \quad j \in J_\delta;$$

then

$$(5.32) \quad m_2 \leq \delta \left( \min_{j \in J_\delta} h_j \right)^{-1} \leq C \frac{\delta^{5/3}}{\mu^{2/3}}, \quad j \in J_\delta.$$

Because of  $\eta_j \geq C\delta$  for  $j \in J_\delta^c$ , it follows from (5.27) that

$$h_j \leq C \left( \frac{\mu}{\delta} \right)^{2/3}, \quad j \in J_\delta^c$$

and hence

$$m_1 \geq (1-\delta) \left( \max_{j \in J_\delta^c} h_j \right)^{-1} \geq C \left( \frac{\delta}{\mu} \right)^{2/3}.$$

Together with (5.32) this proves (5.29) in this case.

Now suppose that in (5.31) we have  $h_j \geq C\mu^{2/5}$ . Then

$$m_2 \leq C\delta\mu^{-2/5}$$

while (5.24) implies that

$$m_1 \geq C\mu^{-2/5}.$$

Together these estimates show once more that (5.29) holds.

Finally consider the case

$$h_j \geq C\mu^{-1/2}, \quad j \in J_\delta.$$

Then

$$m_2 \leq C\delta m_1^{1/2} = C\delta(m_1 + m_2)^{1/2}$$

or

$$m_2^2 - C^2\delta^2 m_2 \leq C^2\delta^2 m_1.$$

By applying Rouché's theorem to the pair of polynomials

$$z^2 - az - am_1 = 0, \quad z^2 - am_1 = 0, \quad a = C^2\delta^2$$

it follows readily that

$$m_2 \leq C\delta m_1^{1/2} \leq C\delta m_1.$$

Thus (5.29) is valid and we have

$$(5.33) \quad m_1 = m(1 + O(\bar{h}^{1/12})).$$

By definition of  $\xi_0$  it follows now from (5.28) and (5.33) that

$$\begin{aligned} \frac{1}{\gamma_0} &= \int_0^1 \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt = \sum_{j \in J_\delta^c} \int_{I_j} \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt + \sum_{j \in J_\delta} \int_{I_j} \left[ \frac{\rho(t)^2}{a(t)} \right]^{1/3} dt \\ &\geq m_1 (12\mu^2)^{1/3} (1 + O(\bar{h}^\epsilon)) = m (12\mu^2)^{1/3} (1 + O(\bar{h}^\epsilon)) \end{aligned}$$

or

$$m\mu^2 \leq \frac{1}{12} \frac{1}{\gamma_0^m} (1+O(\bar{h}^\varepsilon)) .$$

From (5.20) and Theorem 4.5 (modified to the present case) this implies (5.15).

It may be noted that the analogous relation to (4.30), namely,

$$|x_j^\Delta - x_j^0| = O(\bar{h}^\varepsilon)$$

is easily proved when there is only one root of  $u_0''$  in  $\bar{I}$ . In general, the situation appears to be more complicated.

Now we turn to the analog of Theorem 4.8.

Theorem 5.3: For any  $(\lambda, \kappa)$ -regular partition with  $1 \leq \kappa < 2$

$$(5.34) \quad \|e\|_E^2 = \frac{1}{12} \left[ \sum_{j=1}^m \frac{v_j^2 h_j^2}{a_{j-1/2}} \right] (1+O(\bar{h}^\varepsilon))$$

where  $v_j$  is given by (4.44) and  $\varepsilon = 1 - \kappa/2$ .

Proof: Recall that

$$\varepsilon(\Delta)^2 = \frac{1}{12} \sum_{j=1}^m \frac{v_j^2 h_j^2}{a_{j-1/2}} .$$

Using

$$(5.35) \quad v_j^2 = \int_{I_j} (\bar{\rho}_j + \sigma_j(t) + \tau_j(t))^2 dt$$

$$|\sigma_j(x)| \leq Ch_j, \quad \| \tau_j \|_0 \leq C \|e\|_E$$

we obtain, with certain  $|C_i| \leq C$ ,  $i = 1, \dots, 5$ ,

$$\begin{aligned}
 12\varepsilon(\Delta)^2 &= \sum_{j=1}^m \frac{\bar{\rho}_j^2 h_j^3}{a_{j-1/2}} + C_1 \sum_{j=1}^m h_j^5 + C_2 \sum_{j=1}^m I_j \|e\|_E^2 h_j^2 \\
 &+ C_3 \sum_{j=1}^m \bar{\rho}_j h_j^4 + C_4 \sum_{j=1}^m I_j \|e\|_E \bar{\rho}_j h_j^{5/2} + C_5 \sum_{j=1}^m I_j \|e\|_E h_j^{7/2}.
 \end{aligned}$$

By Lemma 5.1 and Theorems 3.1 and 3.2 the first term is asymptotically equal to  $12 \|e\|_E^2$  and from (5.5) it follows that  $I_j \|e\|_E \geq C \bar{h}^K$ . The other terms are then easily estimated to give—with different constants  $C_i$ —

$$\begin{aligned}
 \varepsilon(\Delta)^2 &= \|e\|_E^2 (1 + O(\bar{h}^\varepsilon)) + C_1 \|e\|_E^2 \bar{h}^{4\varepsilon} + C_2 \|e\|_E^2 \bar{h}^2 \\
 &+ C_3 \left[ \sum_{j=1}^m \frac{\bar{\rho}_j^2 h_j^3}{a_{j-1/2}} \right]^{1/2} [\bar{h}^2 + \|e\|_E \bar{h}^{1/2}] + C_4 \|e\|_E^2 \bar{h}^{2\varepsilon+1/2} \\
 &= \|e\|_E^2 (1 + O(\bar{h}^\varepsilon)),
 \end{aligned}$$

which proves (5.34).

Finally, we show that also Theorem 4.9 carries over to this case.

Theorem 5.4: For the partition  $\Delta$  suppose that

$$(5.36) \quad \varepsilon_j = \mu(1 + O(\bar{h}^\varepsilon)) \text{ as } \bar{h}(\Delta) \rightarrow 0, \quad j = 1, \dots, m,$$

where  $\mu$  does not depend on  $j$  and  $\varepsilon = 1/12$ . Then  $\Delta$  is  $(\lambda, 5/3)$ -regular and

$$(5.37) \quad \|e(\Delta)\|_E \leq \|e(\Delta_0)\|_E (1 + O(\bar{h}^\varepsilon)), \text{ as } m \rightarrow \infty.$$

Proof: We show first that  $\Delta$  is  $(\lambda, 5/3)$ -regular. For this note that generally

$$I_j \|z_j\|_E^2 \leq \frac{12}{\pi^2} \epsilon_j^2 (1+O(h_j)), \text{ as } h_j \rightarrow 0, j = 1, \dots, m.$$

This follows from (3.8) in the same manner as (3.14) follows from (3.13b).

Thus by Theorem 3.1 and (5.36) we have

$$I \|e\|_E^2 \leq C m \mu^2,$$

whence by (5.19)

$$\bar{h} \leq C m^{1/5} \mu^{2/5}.$$

Now (5.35) gives

$$(5.38) \quad v_j = I_j \|\bar{\rho}_j + \sigma_j + \tau_j\|_0 \leq C(\bar{\rho}_j h_j^{1/2} + h_j^{3/2} + I_j \|e\|_E)$$

and thus--as in the case of (5.23)--for sufficiently small  $\bar{h}$

$$\underline{h} \geq C \mu^{2/3}.$$

By (5.12) we have

$$\int_{I_j} \rho(x)^2 dx \geq C h_j^3$$

and thus

$$\mu \geq C h_j (I_j \|\rho\|_0 - I_j \|\tau_j\|_0) \geq C(h_j^{5/2} - h_j I_j \|e\|_E)$$

leads to (5.24). The remainder of the proof of the regularity of  $\Delta$  now proceeds exactly as part 3) of the proof of Theorem 5.3.

Similarly the proof of (5.37) follows that of (5.15). In fact, for  $j \in J_6^C$  we have

$$I_j \|\sigma_j\|_0 = \frac{\bar{\rho}_j h_j^{1/2}}{\delta} O(h_j) = \bar{\rho}_j h_j^{1/2} O(\bar{h}^{1-\epsilon})$$

$$I_j \|\tau_j\|_0 = O(\bar{h}) = \frac{\bar{\rho}_j}{\delta} h_j^{1/2} O(\bar{h}^{2\epsilon}) = \bar{\rho}_j h_j^{1/2} O(\bar{h}^\epsilon)$$

which as in (4.47) leads to

$$v_j^2 = \bar{\rho}_j^2 h_j^3 (1 + O(\bar{h}^\epsilon)), \quad j \in J_\delta^c,$$

that is,

$$\mu^2 = \eta_j^2 h_j^3 (1 + O(\bar{h}^\epsilon)), \quad j \in J_\delta^c.$$

Therefore (5.28) holds again. Moreover, because of  $\bar{\rho}_j \leq C\eta_j$  we obtain from (5.38) the estimates (5.30) and (5.31) which in turn imply (5.29).

Now the remaining conclusions of the proof of Theorem 5.4 apply verbatim.

## 6. Numerical Examples

We illustrate the theoretical results with some computational results for the following two sample problems:

### Sample Problem A:

$$(6.1a) \quad -\frac{d}{dx} (x+\alpha)^p \frac{du}{dx} + (x+\alpha)^q u = f, \quad 0 < x < 1, \quad \alpha > 0,$$

$$(6.1b) \quad u(0) = u(1) = 0$$

where  $f$  is chosen such that the solution of (6.1a/b) is

$$(6.1c) \quad u_0(x) = (x+\alpha)^r - [\alpha^r(1-x) + (1+\alpha)^r x], \quad x \in \bar{I}.$$

Here the coefficient functions and  $f$  are analytic in  $\bar{I}$ , and we have  $u_0''(x) \neq 0$  for  $x \in \bar{I}$ . Hence the theory of Section 4 applies. Note that for small  $\alpha$  and negative  $r$  we can create severe near-singularities.

### Sample Problem B:

$$(6.2a) \quad -u'' + u = f, \quad 0 < x < 1,$$

$$(6.2b) \quad u(0) = u(1) = 0$$

where  $f$  is chosen such that the true solution is

$$(6.2c) \quad u_0(x) = e^{\alpha x}(x-\beta) + [\beta(1-x) - e^{\alpha}(1-\beta)x], \quad \alpha \neq 0, \quad \beta = \frac{1}{2} + \frac{2}{\alpha}.$$

Here  $f$  is analytic on  $I$  and  $u_0''(x)$  has a simple root at  $x = 1/2$ , and hence we can apply the theory of Section 5.

The tables of computational results given below include the following data:



$m$	number of intervals used in the partition
$E = 100 \frac{\ e\ _E}{\ u_0\ _E}$	relative error in the energy norm expressed in percent
$E_0 = \frac{1}{12m^3 \gamma_0^2}$	asymptotically smallest error achievable with meshes of $m$ intervals
$\theta = \frac{\ e\ _E}{\varepsilon(\Delta)}$	effectivity quotient (4.39)
$\omega = \left( \max_{j=1, \dots, m} \varepsilon_j \right) / \left( \min_{j=1, \dots, m} \varepsilon_j \right)$	ratio between the largest and smallest value of the error indicators $\varepsilon_j(\Delta)$ of (4.38a)
$x_j$	partition points of the particular mesh

Tables 1 through 6 concern two cases of sample problem A. Each time the left endpoint  $x = 0$  of the interval is near a singularity of  $u_0$  and this is reflected in the fact that the largest and smallest error indicators  $\varepsilon_j(\Delta)$  always occur on the first and last subinterval  $I_1$  and  $I_m$ , respectively. But in the second case the energy expression includes a weight  $(x+\alpha)^2$  which goes strongly down near  $x = 0$ . Thus in this case the near singularity of the solution shows up more weakly under the energy norm.

In all cases the effectivity quotient is less than one and hence the estimate  $\varepsilon(\Delta)$  turns out to be an upper bound of  $\|e\|_E$ . Of course, the theory is only asymptotic in nature and thus  $\varepsilon(\Delta)$  could be smaller than  $\|e\|_E$ . Note that for relative accuracies better than 10% the estimate never overshoots the true error by more than 10%. In fact, for higher accuracies  $\theta$  equals one for all practical purposes. This is in complete agreement with the theory and shows that the a-posteriori error estimate is very reliable

and not at all pessimistic.

In the presence of the near singularity, the use of nonuniform meshes is very advantageous (see, e.g., Table 6), and the approximately optimal meshes produce errors close to the optimal values which by Theorem 4.5 decrease with  $1/m$ . Here the "weaker" singularity of the second case is rather noticeable. The nonuniform meshes are only approximately optimal as the ratio  $\omega$  shows which in each case is reasonably close to one but certainly not equal to it. Nevertheless, as expected, the corresponding errors are clearly not very sensitive to such changes of the mesh except for low accuracies.

Tables 7 through 11 contain results for two cases of sample problem B. Essentially all aspects are the same as for problem A. However, in all cases the maximal  $\epsilon_j$  occurs in the neighborhood of the root  $x_0 = 1/2$  of  $u_0''(x)$ , and, as expected, the ratio  $\omega$  does not converge to one. However, if  $\omega$  is computed only for all intervals outside a small neighborhood of  $x_0$ , then we have again the desired convergence of  $\omega$  to one.

Table 1

Problem A with  $p = 0$ ,  $q = 1$ ,  $r = -1/4$ ,  $\alpha = 1/100$   
 Uniform mesh,  ${}_I\|u_0\|_E = 6.09811$

m	E	$E_0$	$\theta$	$\omega$
5	85.301	22.613	.1706	8.84(+6)
10	73.768	11.306	.2950	2.34(+7)
20	58.784	5.653	.4702	5.31(+7)
40	41.933	2.827	.6708	1.11(+8)
80	26.316	1.413	.8419	2.19(+8)

Table 2

Problem A with  $p = 0$ ,  $q = 1$ ,  $r = -1/4$ ,  $\alpha = 1/100$   
 Approximately optimal mesh,  ${}_I\|u_0\|_E = 6.09811$

m	E	$E_0$	$\theta$	$\omega$
5	22.243	22.613*	.6524	5.854
10	11.289	11.306	.9025	2.274
20	5.652	5.653	.9757	1.372
40	2.826	2.827	.9940	1.111
80	1.413	1.413	.9984	1.031

\*Note here the asymptotic nature of the estimate  $E_0$  of the lowest achievable error.

Table 3

Problem A with  $p = 2$ ,  $q = 1$ ,  $r = -1$ ,  $\alpha = 1/100$   
 Uniform mesh,  $\|u_0\|_E = 0.28678$

m	E	$E_0$	$\theta$	$\omega$
5	25.215	9.619	.3270	3.227(+3)
10	13.696	4.809	.4467	4.640(+3)
20	7.296	2.404	.5966	8.316(+3)
40	3.813	1.202	.7610	1.055(+4)
80	1.959	0.601	.8947	1.308(+4)

Table 4

Problem A with  $p = 2$ ,  $q = 1$ ,  $r = -1$ ,  $\alpha = 1/100$   
 Approximately optimal mesh,  $\|u_0\|_E = 0.286$

m	E	$E_0$	$\theta$	$\omega$
5	9.994	9.619	.7679	3.355
10	4.809	4.809	.9331	1.666
20	2.409	2.404	.9828	1.184
40	1.203	1.202	.9956	1.049
80	0.601	0.601	.9989	1.012

Table 5

Problem A: Approximately optimal mesh for  $m = 10$

Case A1:  $p = 0, q = 1, r = -1/4, \alpha = 1/100$

Case A2:  $p = 2, q = 1, r = -1/4, \alpha = 1/100$

j	$x_j$	
	Case A1	Case A2
0	.0000	.0000
1	.00207	.0127
2	.00487	.0340
3	.00877	.0676
4	.01443	.1170
5	.02308	.1862
6	.03732	.2798
7	.06318	.4025
8	.11781	.5598
9	.26831	.7569
10	1.00000	1.0000

Table 6

Problem A with  $p = 0$ ,  $q = 1$ ,  $r = -1/4$ ,  $\alpha = 1/100$   
 Partitions obtained from the uniform mesh by successively subdividing  
 the first interval into half

m	E	$E_0$	$\theta$	$\omega$
10	73.768	11.306	.2950	2.34(+7)
11	58.853	10.278	.4705	5.85(+6)
12	42.286	9.421	.6922	1.45(+6)
13	27.357	8.696	.8457	3.45(+5)
14	17.634	8.075	.9340	7.49(+4)
15	13.580	7.530	.9447	1.41(+4)

Table 7

Problem B with  $\alpha = 1$ ,  $\beta = 5/2$   
Uniform mesh,  $\|u_0\|_E = .071070$

m	E	$E_0$	$\theta$	$\omega$
5	43.462	32.317	.9759	1.126(+2)
10	22.080	16.158	.9939	1.757(+2)
20	11.083	8.079	.9984	7.568(+2)
40	5.547	4.039	.99924	3.142(+3)
80	2.774	2.019	.99990	1.281(+4)

Table 8

Problem B with  $\alpha = 1$ ,  $\beta = 5/2$   
Approximately optimal mesh,  $\|u_0\|_E = .071070$

m	E	$E_0$	$\theta$	$\omega$
5	33.869	32.317	.9466	1.577
10	16.519	16.158	.9694	1.676
20	8.153	8.079	.9823	1.755
40	4.049	4.039	.9894	1.788
80	2.018	2.019	.9933	2.437

Table 9

Problem B with  $\alpha = 5$ ,  $\beta = 9/10$   
Uniform mesh

m	E	$E_0$	$\theta$	$\omega$
5	49.477	18.174	.9059	4.049(+3)
10	26.554	9.087	.9742	1.229(+4)
20	13.530	4.543	.9934	4.621(+4)
40	6.797	2.271	.9983	1.808(+5)
80	3.403	1.135	.9995	7.173(+5)

Table 10

Problem B with  $\alpha = 5$ ,  $\beta = 9/10$   
Approximately optimal mesh

m	E	$E_0$	$\theta$	$\omega$
5	17.021	18.174	.7988	2.617
10	9.181	9.087	.9217	2.822
20	4.521	4.543	.9595	2.324
40	2.254	2.271	.9820	1.661
80	1.138	1.135	.9958	1.614



Table 9

Problem B with  $\alpha = 5$ ,  $\beta = 9/10$   
Uniform mesh

m	E	$E_0$	$\theta$	$\omega$
5	49.477	18.174	.9059	4.049(+3)
10	26.554	9.087	.9742	1.229(+4)
20	13.530	4.543	.9934	4.621(+4)
40	6.797	2.271	.9983	1.808(+5)
80	3.403	1.135	.9995	7.173(+5)

Table 10

Problem B with  $\alpha = 5$ ,  $\beta = 9/10$   
Approximately optimal mesh

m	E	$E_0$	$\theta$	$\omega$
5	17.021	18.174	.7988	2.617
10	9.181	9.087	.9217	2.822
20	4.521	4.543	.9595	2.324
40	2.254	2.271	.9820	1.661
80	1.138	1.135	.9958	1.614

Table 11

Problem B: Approximately optimal mesh for  $m = 10$

Case B1:  $\alpha = 1, \beta = 5/2$

Case B2:  $\alpha = 5, \beta = 9/10$

j	$x_j$	
	Case B1	Case B2
0	.0000	.0000
1	.0097	.4192
2	.1859	.6918
3	.3001	.7715
4	.5218	.8255
5	.6872	.8673
6	.7754	.9016
7	.8442	.9309
8	.9025	.9565
9	.9538	.9794
10	1.0000	1.0000

### References

1. N. I. Akhiezer, The Calculus of Variations, Translated from the Russian by A. H. Frink, Blaisdell Publishing Co., New York, 1962.
2. I. Babuška, The self-adaptive approach in the finite element method, in The Mathematics of Finite Elements and Applications II, MAFELAP 1975, ed. J. R. Whiteman, Academic Press, London, 1976, 125-142.
3. I. Babuška, W. Rheinboldt, A-posteriori error estimates for the finite element method, University of Maryland, Computer Science Technical Report TR-581, 1977; Int. J. Num. Meth. Eng., accepted.
4. I. Babuška, W. Rheinboldt, Error estimates for adaptive finite element computations, University of Maryland, Computer Science Technical Report BN-854, 1977; SIAM J. Num. Anal., in press.
5. I. Babuška, W. Rheinboldt, Computational aspects of finite element analysis, in Mathematical Software - III, ed. J. R. Rice, Academic Press, New York, 1977, 223-253.
6. I. Babuška, W. Rheinboldt, C. Mesztenyi, Self-adaptive refinement in the finite element method, University of Maryland, Computer Science Technical Report TR-375, 1975.
7. H. G. Burchard, Splines (with optimal knots) are better, J. Appl. Anal. 3, 1974, 309-319.
8. H. G. Burchard, D. F. Hale, Piecewise polynomial approximations on optimal meshes, J. Approx. Theory 14, 1975, 128-147.
9. W. E. Carroll, R. M. Barker, A theorem for optimum finite element idealization, Inst. J. Solids Structures 9, 1973, 883-895.
10. C. deBoor, Good approximation by splines with variable knots, in Spline Functions and Approximation Theory, ed. A. Meir, A. Sharma, ISNM Vol. 21, Birkhäuser Verlag, Basel, 1973, 57-72.
11. C. deBoor, Good approximations with variable knots II, Lecture Notes in Mathematics, Vol. 363, Springer Verlag, Germany, 1973, 12-20.
12. C. deBoor, J. R. Rice, An adaptive algorithm for multivariate approximation giving optimal convergence rates, University of Wisconsin, Mathematics Research Center, MRC Technical Summary Report 1773, 1977.
13. C. deBoor, B. Swartz, Collocation at Gaussian points, SIAM J. Num. Anal. 10, 1973,

14. V. E. Denny, R. B. Landis, A new method for solving two-point boundary value problems using optimal node distributions, J. Comp. Phys. 9, 1972, 120-137.
15. T. E. Hull, Numerical solution of initial value problems for ordinary differential equations, in Numerical Solutions of Boundary Value Problems for Ordinary Differential Equations, ed. A. K. Aziz, Academic Press, New York, 1975, 3-26.
16. H. B. Keller, Numerical solution of boundary value problems for ordinary differential equations: Survey and some recent results on difference methods, in Numerical Solution of Boundary Value Problems for Ordinary Differential Equations, ed. A. K. Aziz, Academic Press, New York, 1975, 27-88.
17. M. Lentini, V. Pereyra, Boundary problem solvers for first-order systems based on deferred corrections, in Numerical Solution of Boundary Value Problems for Ordinary Differential Equations, ed. A. K. Aziz, Academic Press, New York, 1975, 293-316.
18. R. J. Melosh, D. E. Killian, Finite element analysis to attain a pre-specified accuracy, Preprint, presented at 2nd National Symposium on Computerized Structural Analysis, George Washington University, Washington, D.C., 1976.
19. R. J. Melosh, P. V. Marcal, An energy basis for mesh refinement of structural continua, Int. J. Num. Meth. Eng. 11, 1977, 1083-1091.
20. J. A. Nitsche, A. H. Schatz, Interior estimates for Ritz-Galerkin methods, Math. Comp. 28, 1974, 937-948.
21. G. M. Phillips, Error estimates for best approximation, in Approximation Theory, ed. A. Talbot, Academic Press, London, 1970, 1-6.
22. C. E. Pearson, On a differential equation of boundary layer type, J. Math. Phys. 47, 1968, 134-154.
23. G. Sewell, An adaptive computer program for the solution of  $\text{Div}(p(x,y)\text{grad } u) = f(x,y,u)$  on a polygonal region, in The Mathematics of Finite Elements and Applications II, MAFELAP 1975, ed. J. Whiteman, Academic Press, London, 1976, 125-144.
24. G. Strang, G. J. Fix, An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs, NJ, 1973.

25. D. J. Turcke, Further developments in grid selection procedures in the finite element method, in Proceedings of the Fifth Canadian Congress of Applied Mechanics, University of New Brunswick, Fredericton, N.B. Canada, 1975.
26. D. J. Turcke, G. M. McNeice, A variational approach to grid optimization in the finite element method, in Conference on Variational Methods in Engineering, Southampton University, England, 1972.
27. D. J. Turcke, G. M. McNeice, Guidelines for selecting finite element grids based on an optimization study, Computers and Structures 4, 1973, 499-519.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ONR-N00014-77-C-0623-869	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ANALYSIS OF OPTIMAL FINITE ELEMENT MESHES IN $R^1$		5. TYPE OF REPORT & PERIOD COVERED Technical Report
7. AUTHOR(s) I. Babuska W. C. Rheinboldt		6. PERFORMING ORG. REPORT NUMBER BN-8691
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Physical Science & Technology University of Maryland College Park, MD 20742		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0623 E(401)3443
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE 11 Mar 1978
		13. NUMBER OF PAGES 55 1263 p.
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Abstract continued ...numerical experiments show the results to be excellent already for 10% accuracy. The approaches are not restricted to the model problem considered here only for clarity; in fact, they allow for rather straightforward extensions to more general problems in one dimension as well as to higher order elements.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) finite element method error estimates optimal mesh mesh refinement		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A theory of a-posteriori estimates for the finite element method was developed earlier, by the authors. Based on this theory, for a two-point boundary value problem the existence of a unique optimal mesh distribution is proved and its properties analyzed. This mesh is characterized in terms of certain, easily computable local error indicators which in turn allow for a simple adaptive construction of the mesh and also permit the computation of a very effective a-posteriori error bound. While the error estimates are asymptotic in nature,		

DD FORM 1473  
1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410 140